



KARAKTERISTIK BUTIR SOAL: CLASSICAL TEST THEORY VS ITEM RESPONSE THEORY?

Muh. Syahrul Sarea¹, Rosnia Ruslan²

sareasyahrul@gmail.com¹

Prodi Pendidikan Agama Islam IAIN Bone¹, Dinas Pendidikan Kab. Banjar²

Abstract

This research aims to describe the characteristic of UAS items theme 1 at the fourth grade of Primary School in Paramasan bawah village according to the item difficulties and discrimination. The sample of this research was 37 students who took the final examination year academic 2018/2019. The objects of this research were question items and the answer sheet of the final exam that obtained from 3 different schools in Paramasan Bawah village. The data analysis technique used in this research was empirical analysis helped by Bilog and Iteman program application. This analysis used to know the characteristic of items based on the Item Response Theory and Classical Test Theory. The result of this research showed the characteristic of UAS items, according to item response theory, 30 items had a good discrimination and 33 items had a good item difficulty, while according to Classical Test Theory: 15 items had a good discrimination and 27 items had a good item difficulty.

Keywords: *Characteristics of items, item difficulties, discrimination*

Diterima: 12-07-2019, Direview: 07-08-2019, Diterbitkan: 15-08-2019

PENDAHULUAN

Kegiatan evaluasi selalu menjadi bagian yang sangat krusial untuk menentukan apakah aktivitas yang telah dilakukan berhasil atau tidak (Mansyur & Harun, 2015). Evaluasi merupakan salah satu rangkaian kegiatan dalam meningkatkan kualitas, kinerja atau produktivitas suatu lembaga dalam melaksanakan programnya (Mardapi, 2012). Pengukuran merupakan suatu langkah atau tindakan yang harus diambil untuk melaksanakan evaluasi. Dikatakan harus diambil karena semua tindakan evaluasi harus didahului oleh kegiatan pengukuran. Pengukuran merupakan proses pemberian angka yang

diharapkan dapat menunjukkan kemampuan peserta didik mengenai suatu mata pelajaran (Mardapi, 2012). Oleh karena itu didalam praktik, masalah pengukuran mempunyai kedudukan sangat penting didalam proses evaluasi.

Pada proses pengukuran, dibutuhkan alat ukur. Alat ukur inilah yang memberikan informasi tentang posisi seseorang dalam atribut yang diukur sehingga, untuk memperoleh hasil pengukuran yang dapat menggambarkan hasil pengukuran sebenarnya dibutuhkan alat ukur dengan tingkat validitas dan reliabilitas yang tinggi. Sejalan dengan itu, (Saifuddin, 2002) mengatakan bahwa instrumen yang baik adalah instrumen yang mampu menghasilkan data dan memberikan informasi yang akurat agar informasi yang diperoleh dari hasil pengukuran menggambarkan kemampuan peserta didik yang sebenarnya.

Secara umum dalam mengukur prestasi belajar siswa digunakan instrument tes. (Allen & Yen, 1979) menyebut tes sebagai *“a test is device for optaining a sample of an indifidual’s behavior”*. Hal yang hampir sama juga dikemukakan oleh (Mardapi, 2012) mengatakan bahwa tes merupakan salah satu bentuk instrumen yang digunakan untuk melakukan pengukuran, yaitu mengumpulkan informasi karakteristik suatu objek. Ahli pengukuran yang lain, (Widoyoko, 2012) mengatakan bahwa tes dapat diartikan sebagai sejumlah pertanyaan yang harus diberikan tanggapan dengan tujuan untuk mengukur tingkat kemampuan seseorang atau mengungkap aspek tertentu dari orang yang dikenai tes”.

Soal ujian akhir semester merupakan salah satu bentuk instrument tes yang mengukur pencapaian pembelajaran di sekolah baik terhadap peserta didik maupun terhadap guru. Hasil ujian akhir semester menggambarkan pencapaian ketuntasan standar kompetensi peserta didik dan kualitas pembelajaran yang diterapkan oleh guru sehingga instumen tes yang digunakan harus memiliki karakteristik butir yang baik serta representatif dalam mengukur setiap aspek pencapaian peserta didik yang sebenarnya.

Dalam membuat soal ujian akhir semester bukanlah perkara mudah bagi guru mengingat tingkat pengetahuan guru dalam membuat dan mengembangkan soal sangat beragam. Diperlukan analisis butir soal dengan tingkat validitas dan reliabilitas yang baik, sehingga komposisi soal mudah, sedang dan sukar menyebar secara proporsional sesuai dengan materi pelajaran yang diujikan.

Soal yang baik akan dapat menyeleksi secara alamiah peserta didik yang cerdas dan peserta didik kurang cerdas melalui hasil evaluasi belajar peserta didik. Hasil Penelitian yang dilakukan oleh (Sarea & Hadi, 2015) mengatakan bahwa terdapat sekolah dengan kualitas soal ujian akhir semester (UAS) yang baik sebanyak 7,14%. Hal ini

menggambarkan bahwa dalam pembuatan soal ujian akhir semester (UAS) bukanlah hal yang mudah sehingga perlu untuk diketahui karakteristiknya.

Karakteristik butir soal meliputi tingkat kesukaran butir soal yang baik dengan komposisi soal mudah, sedang dan sukar menyebar secara proporsional sesuai dengan materi pelajaran yang diujikan dan indeks daya beda butir soal yang baik mampu untuk membedakan antara kelompok peserta didik berkemampuan tinggi dengan kelompok peserta didik berkemampuan rendah, sehingga hasil evaluasi belajar peserta didik akan menggambarkan hasil belajar peserta didik yang sebenarnya. Terdapat 2 pendekatan yang dapat digunakan untuk melihat karakteristik butir soal yaitu *Classical Test Theory (CTT)* dan *Item Response Theory (IRT)*.

Teori Tes Klasik (*Classical Test Theory*)

Salah satu teori pengukuran yang tertua di dunia pengukuran behavioral adalah *classical true-score theory*. Teori ini dalam bahasa Indonesia sering disebut dengan teori tes klasik. Teori tes klasik merupakan sebuah teori yang mudah dalam penerapannya serta model yang cukup berguna dalam mendeskripsikan bagaimana kesalahan dalam pengukuran dapat mempengaruhi skor amatan.

Teori tes klasik merupakan upaya menjelaskan error pada pengukuran. Pada teori tes klasik, model error pengukuran berdasarkan koefisien korelasi. Koefisien korelasi ditemukan oleh Charles Spearman, merupakan upaya menjelaskan error menggunakan dua komponen: korelasi sebenarnya dan korelasi amatan.

Koefisien korelasi dalam teori tes klasik, berdasarkan pada teori bahwa nilai rata-rata pengukuran dari semua hasil pengukuran yang mungkin akan sama pengukuran sebenarnya pada populasi. Akibatnya terhadap teori tes klasik adalah: 1) error bersifat acak dan 2) pengukuran terdiri atas tiga komponen: yakni indikator amatan, indikator hipotetikal yang menunjukkan nilai populasi murni, dan konsep hipotetikal yang menunjukkan kuantitas ketidak sesuaian antara indikator sebenarnya dan indikator amatan. Teori tes klasik dinyatakan dengan rumus $X = T + E$. Menurut (Mardapi, 2012) pada teori tes klasik, skor yang tampak (X) terdiri atas skor sebenarnya atau skor murni (T) dan skor kesalahan (E).

Teori tes klasik menekankan pada skor mentah dari satu ujian yang dihasilkan. Skor mentah menunjukkan kemampuan seseorang. Dari skor mentah ini maka berbagai analisis dan interpretasi bisa dihasilkan sesuai dengan keperluan studi yang dilakukan (Sumintono & Widhiarso, 2014).

Menurut (Hambleton & Jones, 1993) “*Classical test models are often referred to as "weak models" because the assumptions of these models are fairly easily met by test data*”. Teori tes klasik mempunyai beberapa asumsi. Pertama, instrumen hanya satu dimensi. Kedua, skor kesalahan pengukuran tidak berinteraksi dengan skor sebenarnya. Ketiga, skor kesalahan tidak berkorelasi dengan skor sebenarnya dan skor-skor kesalahan pada tes yang lain untuk peserta tes yang sama. Keempat, rata-rata skor kesalahan sama dengan nol. Asumsi tersebut dijadikan dasar untuk mengembangkan rumus yang digunakan untuk menghitung reliabilitas tes.

Reliabilitas dapat diartikan sebagai keajegan atau konsistensi hasil pengukuran atau hasil tes yang dilakukan pada waktu yang berbeda pada subjek yang sama. (Allen & Yen, 1979) menyatakan bahwa tes disebut reliabel jika skor amatan mempunyai korelasi yang tinggi dengan skor yang sebenarnya. Reliabilitas merupakan koefisien korelasi antara dua skor amatan yang diperoleh dari hasil pengukuran menggunakan tes yang paralel.

Penentuan harga reliabilitas suatu tes dapat digunakan beberapa cara dan formula. Formula yang sering digunakan adalah Formula belah dua, alpha Cronbach, Guttman, Flanagan, Raju, dan paralel. Nilai hasil perhitungan dari formula tersebut sering dikatakan sebagai koefisien reliabilitas. Besarnya indeks keandalan yang diterima adalah minimal 0,70. Besarnya indeks reliabilitas menyatakan besar kesalahan pengukuran. Semakin besar indeks reliabilitas maka akan semakin kecil kesalahan pengukuran, demikian sebaliknya (Mardapi, 2012)

Sayangnya, seperti diungkapkan (Saifuddin, 2002) bahwa teori tes klasik memiliki keterbatasan yang mendasar yaitu hasil estimasi parameter tergantung pada karakteristik peserta ujian (*group dependent*). Hal ini berimplikasi pada tingkat kesukaran soal akan menjadi rendah jika tes diujikan pada kelompok peserta tes berkemampuan tinggi dan sebaliknya jika tes diujikan pada peserta dengan kemampuan rendah maka tingkat kesukaran tes itu akan tinggi. Kedua hasil estimasi kemampuan peserta tergantung pada karakteristik butir soal (*item dependent*). Keterbatasan ini menyebabkan estimasi kemampuan peserta akan rendah jika soal yang diberikan berada di atas kemampuannya. Sebaliknya, estimasi kemampuan peserta akan tinggi bila soal yang diujikan berada dibawah tingkat kemampuannya.

Mengatasi kelemahan pada teori tes klasik, maka para ahli pengukuran mengembangkan model yang tidak terikat dengan sampel (*sample free*). Model ini

selanjutnya dikenal tes modern atau tes respon butir. Menurut teori respon butir, perilaku seseorang dapat dijelaskan oleh karakteristik orang yang bersangkutan sampai pada batas-batas tertentu (Mardapi, 2012)

Teori Respon Butir (*Item Response Theory*)

Van der Linden & Hambleton (2013) menyatakan bahwa teori respon butir (IRT) merupakan salah satu cara untuk menilai kelayakan butir dengan membandingkan rerata penampilan butir terhadap tampilan bukti kemampuan kelompok yang diramalkan oleh model. (Liang, Wells, & Hambleton, 2014) mengatakan bahwa “*Item response theory (IRT) is a powerful scaling technique with appealing features such as the invariance of item and ability parameter values*”. Tujuan utama teori respon butir dikembangkan adalah untuk mengatasi teori tes klasik yang tidak *independent* terhadap kelompok peserta yang mengerjakan tes maupun terhadap tes yang diujikan.

Bagian penting dari teori respon butir adalah probabilitas jawaban benar peserta tes, parameter butir dan parameter peserta tes dihubungkan melalui suatu fungsi matematik atau model formula matematik. Dalam formula ini, nilai kemungkinan peserta tes menjawab soal dipahami sebagai fungsi logistik perbedaan parameter yang dimasukkan kedalam model.

Teori respon butir atau teori tes modern dikembangkan atas dasar pikiran: 1) hasil ujian seseorang dapat diprediksi dari kemampuan yang dimilikinya dan 2) hubungan antara hasil ujian dan kemampuan dinyatakan dalam sebuah fungsi yang disebut kurva karakteristik butir atau *Item Characteristic Curve* (Hambleton, Swaminathan, & Rogers, 1991). Fungsi kurva karakteristik butir (ICC) ini memperlihatkan posisi peserta tes dengan kemampuan tinggi akan memiliki peluang yang lebih sebaliknya, peserta tes dengan kemampuan rendah besar menjawab butir soal dengan tingkat kesukaran yang tinggi. Kemampuan ini sering disebut dengan potensi yang merupakan faktor dominan untuk menentukan keberhasilan seseorang dalam belajar yang ditunjukkan oleh hasil yang diperoleh dari suatu ujian.

Hubungan dengan tingkat kesukaran butir soal, sifat *invariants* berarti bahwa indeks kesukaran suatu butir soal tidak akan berubah, sekalipun soal dikerjakan oleh peserta tes yang pandai atau kurang pandai. Kondisi tersebut tidak berlaku bagi tes klasik sehingga sifat *invariants* merupakan salah satu kelebihan teori respon butir. Teori tes modern atau respon butir memiliki syarat yang lebih ketat dibanding tes klasik, baik ditinjau dari asumsinya maupun ukuran sampel yang dibutuhkan dalam analisisnya.

Asumsi Teori Respon Butir

Dalam teori respon butir, model matematisnya mempunyai makna bahwa probabilitas subjek untuk menjawab butir dengan benar tergantung pada kemampuan subjek dan karakteristik butir. Ini berarti bahwa peserta tes dengan kemampuan tinggi akan mempunyai probabilitas menjawab benar lebih besar jika dibandingkan dengan peserta yang mempunyai kemampuan rendah. Hambleton et al. (1991) menyatakan bahwa ada tiga asumsi yang mendasari teori respon butir, yaitu *unidimensi*, *independensi lokal* dan invariansi parameter.

1. *Unidimensi*, artinya setiap butir tes hanya mengukur satu kemampuan. Contohnya, pada tes prestasi belajar bidang studi matematika, butir-butir yang termuat di dalamnya hanya mengukur kemampuan peserta didik dalam bidang studi matematika saja, bukan bidang yang lainnya. Pada praktiknya, asumsi *unidimensi* sulit dilakukan karena adanya faktor-faktor kognitif, kepribadian dan faktor-faktor pelaksanaan tes, seperti kecemasan, motivasi, dan tendensi untuk menebak. Oleh karena itu, asumsi *unidimensi* dapat ditunjukkan hanya jika tes mengandung satu komponen dominan yang mengukur prestasi subjek. Menurut Naga (1992) Salah satu cara menguji ketercapaian syarat unidimensi adalah metode analisis faktor
2. *Independensi lokal* menyatakan bahwa sikap kemampuan yang mempengaruhi suatu tes adalah konstan, maka respon peserta tes pada setiap butir soal adalah independen secara statistik. Sebagaimana dijelaskan oleh van der Linden & Hambleton (2013) :

The term “local” in the local independence assumption is used to indicate that responses are assumed independent at the level of individual persons with the same value of, but the assumption does not generalize to the case of variation in.

Asumsi independensi lokal menyatakan bahwa tidak ada korelasi antara respon peserta tes pada butir soal yang berbeda. Hal ini menunjukkan bahwa kemampuan yang dinyatakan dalam model adalah satu-satunya faktor yang mempengaruhi respon peserta tes pada butir-butir soal. Bukti yang lain adalah peluang dari pola jawaban setiap peserta tes sama dengan hasil kali peluang jawaban peserta tes pada setiap butir soal.

3. *Invariansi parameter* artinya bahwa karakteristik butir soal tidak tergantung pada distribusi parameter kemampuan peserta tes dan parameter yang menjadi ciri peserta tes tidak bergantung dari ciri butir soal. Schulz & Fraillon (2011) mengatakan “*Measurement invariance holds if individuals with the same score on the same*

measurement instrumen have the same standing on the underlying construct that is measured". Kemampuan seseorang tidak akan berubah hanya karena mengerjakan tes yang berbeda tingkat kesulitannya dan parameter butir tes tidak akan berubah hanya karena diujikan pada kelompok peserta tes yang berbeda tingkat kemampuannya.

METODE PENELITIAN

Penelitian ini merupakan jenis penelitian deskriptif kuantitatif dengan data berupa perangkat soal UAS dan lembar jawaban peserta didik Kelas IV Tema 1 Sekolah Dasar. Penelitian ini bertujuan untuk mengetahui karakteristik soal UAS Kelas IV Tema 1 Sekolah Dasar tahun pelajaran 2018/2019 di Desa Paramasan Bawah. Respon peserta didik dianalisis secara empiris berdasarkan pendekatan teori tes klasik dan teori respon butir untuk menguji kelayakan butir soal. Penelitian dilaksanakan pada bulan november sampai dengan Januari 2019 di desa paramasan bawah Kabupaten Banjar Provinsi Kalimantan Selatan

Subjek penelitian ini adalah 37 peserta didik yang mengikuti UAS semester gasal tema satu tingkat sekolah dasar (SD) tahun pelajaran 2018/2019 di desa paramasan bawah. Objek penelitian ini adalah perangkat soal dan seluruh lembar jawaban UAS yang diperoleh dari 3 sekolah di desa paramasan bawah. Lembar jawaban peserta didik akan dianalisis untuk melihat karakteristik soal UAS semester gasal tema satu kelas IV Sekolah dasar (SD) di desa paramasan bawah tahun pelajaran 2018/2019 berdasarkan tabel 1

Tabel 1. Daftar Jumlah Peserta Didik dan Sekolah

Nama Sekolah	Peserta Didik
SDN Paramasan bawah 1	12
SDN Paramasan bawah 2	17
SDN Paramasan bawah 3	9

Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini dilakukan dengan instrumen dokumen. Dokumen yang digunakan untuk memperoleh data dalam penelitian ini antara lain: 1) Spesifikasi soal UAS Kelas IV pada tema satu tingkat sekolah dasar (SD), 2) lembar soal UAS Kelas IV pada tema satu tingkat Sekolah Dasar (SD) dan 3) Lembar jawaban peserta didik pada tema satu kelas IV tingkat Sekolah Dasar (SD) tahun pelajaran 2018/2019 di Desa Paramasan Bawah.

Teknik Analisis Data

Data yang diperoleh dalam penelitian ini, selanjutnya dianalisis secara empiris. Respon peserta didik terhadap perangkat tes dianalisis dengan menggunakan paket program Bilog dan iteman.

Analisis dengan Pendekatan *Item Response Theory* menggunakan Program Bilog

Ada tiga model logistik dalam teori respon butir, yaitu model logistik satu parameter, model logistik dua parameter, dan model logistik tiga parameter. Perbedaan dari ketiga model tersebut pada banyaknya parameter yang digunakan dalam menggambarkan karakteristik butir dalam model yang digunakan.

Model parameter logistik yang cocok dengan instrumen dapat dilihat dari kecenderungan butir instrumen UAS cocok menggunakan model logistik 1 parameter (1 PL), model logistik 2 parameter (2 PL), atau model logistik 3 parameter (3 PL) dilihat dari fit model ketiga parameter logistik tersebut. Menurut Retnawati (2014) pemilihan model parameter logistik dilihat dari jumlah butir yang cocok (Fit Model) paling banyak sebagai model untuk analisis data.

Analisis dengan *Bilog* digunakan untuk menganalisis butir soal. Analisis butir ini menghasilkan output dalam 3 fase. Fase pertama akan mendeskripsikan tentang valid atau tidak validnya setiap butir tes yang diberikan kepada peserta didik, fase kedua akan menghasilkan karakteristik dari setiap butir soal yang dikerjakan oleh peserta didik baik tingkat kesukaran dan daya pembeda sedangkan fase ketiga yaitu deskripsi kemampuan peserta didik dari yang paling rendah hingga yang paling tinggi.

1. Berdasarkan teori respon butir khususnya yang menggunakan program *Bilog*, butir soal yang baik memiliki indeks kesukaran berkisar -2 sampai $+2$. indeks kesukaran mendekati -2 dikategorikan sebagai soal mudah, sedangkan indeks kesukaran mendekati $+2$ dikategorikan sebagai soal yang sukar (Hambleton et al., 1991).
2. Berdasarkan teori respon butir khususnya yang menggunakan program *Bilog*, butir soal dengan indeks daya pembeda secara teoritis memiliki skala $(-\infty$ sampai $+\infty)$. Daya pembeda 0 tidak baik, karena kondisi ini menyebabkan $P_i(\theta)$ menjadi konstan dan membentuk garis lurus. Daya pembeda negatif juga tidak baik hal ini menyebabkan kemiringan daya pembeda negatif, sehingga kurva yang terbentuk bukan monoton naik. Sarea (2018) menyatakan bahwa indeks daya beda antara 0,4 sampai dengan -2 dapat dikategorikan baik.

Analisis dengan Pendekatan *Classical Test Theory* menggunakan Program *Iteman*

Hasil dari analisis butir soal mencakup informasi mengenai (1) parameter tingkat kesukaran butir soal dan (2) indeks daya pembeda butir soal.

1. Parameter tingkat kesukaran butir soal (p) adalah proporsi peserta yang menjawab benar butir soal. Besarnya indeks kesukaran butir soal yang diterima untuk menyatakan butir soal dikatakan baik adalah 0,3 sampai 0,8 dilihat dari nilai yang ada pada *prop Correct* (Mardapi, 2012).
2. *Indeks* Daya pembeda butir soal adalah besarnya korelasi antara rata-rata skor peserta yang menjawab benar pada butir soal dengan rata-rata skor total. Hal ini dihitung dengan melihat besarnya nilai koefisien *point biserial* (p_{pbis}). Mardapi (2012) mengatakan bahwa besarnya daya pembeda mulai dari -0,1 sampai +0,1. Perhitungan indeks daya pembeda butir soal dapat dilakukan dengan beberapa cara antara lain, menggunakan indeks korelasi, indeks keselarasan ataupun indeks diskriminasi. Besarnya indeks daya pembeda minimal 0,3.

HASIL PENELITIAN DAN PEMBAHASAN

Karakteristik Butir Soal Berdasarkan Teori Respon Butir

Pada analisis menggunakan program *bilog*, sebelum, menganalisis karakteristik butir soal baik tingkat kesukaran butir maupun daya pembeda butir soal terlebih dahulu dilakukan analisis kecocokan butir dengan analisis model logistik yang digunakan sehingga jumlah butir soal yang dianalisis lebih akurat. Menurut (Retnawati, 2014) butir yang memiliki kecocokan paling banyak dengan model digunakan sebagai model untuk analisis data. Berdasarkan hasil analisis data diperoleh bahwa model yang cocok digunakan untuk menganalisis data menggunakan model 2 parameter logistik dimana tingkat kecocokan butir dengan model sebanyak 37 dari 41 butir yang di analisis.

Berdasarkan hasil analisis data menggunakan program Bilog Model 2 Parameter logistik meninjau karakteristik soal UAS Kelas IV Tema satu Sekolah Dasar pada aspek tingkat kesukaran dan daya beda yang memiliki kriteria yang baik seperti pada tabel 2.

Tabel 2. Kriteria tingkat kesukaran dan indeks daya beda yang baik

Parameter/ Prob	Nilai	Keterangan
a	0,4 s/d 2	Baik
b	-2 s/d 2	Baik
Prob	0,05	Fit Model

Berdasarkan kriteria di atas maka diperoleh data hasil analisis menggunakan program Bilog pada model 2 parameter seperti pada tabel 3 berikut.

Tabel 3. Karakteristik butir Soal Berdasarkan Teori Respon Butir Model 2 Parameter

Butir	Daya Beda		Tingkat Kesukaran		Fit Model
1	0.427	Baik	2.568	Tidak Baik	Fit model
3	0.684	Baik	1.249	Baik	Fit model
5	0.355	Tidak Baik	0.752	Baik	Fit model
6	0.300	Tidak Baik	2.140	Tidak Baik	Fit model
8	0.761	Baik	0.236	Baik	Fit model
9	0.340	Tidak Baik	3.505	Tidak Baik	Tidak Fit model
10	0.303	Tidak Baik	3.441	Tidak Baik	Tidak Fit model
11	0.428	Baik	-1.393	Baik	Fit model
14	0.314	Tidak Baik	2.637	Tidak Baik	Tidak Fit model
15	0.418	Baik	0.304	Baik	Fit model
16	0.340	Tidak Baik	2.461	Tidak Baik	Fit model
18	0.300	Tidak Baik	1.516	Baik	Fit model
19	0.320	Tidak Baik	2.257	Tidak Baik	Fit model
21	0.465	Baik	0.427	Baik	Fit model
22	0.504	Baik	0.394	Baik	Fit model
23	0.623	Baik	0.717	Baik	Fit model
24	0.926	Baik	-0.665	Baik	Fit model
25	0.427	Baik	0.634	Baik	Fit model
26	0.724	Baik	0.510	Baik	Fit model
27	0.535	Baik	0.770	Baik	Fit model
28	0.382	Tidak Baik	1.505	Baik	Fit model
30	0.620	Baik	-0.300	Baik	Fit model
31	1.156	Baik	-0.211	Baik	Tidak Fit model
32	0.484	Baik	0.111	Baik	Fit model
33	0.417	Baik	1.825	Baik	Fit model
34	0.441	Baik	0.616	Baik	Fit model
35	0.509	Baik	0.245	Baik	Fit model
36	0.455	Baik	-0.657	Baik	Fit model
37	0.584	Baik	-0.565	Baik	Fit model
39	0.498	Baik	-0.472	Baik	Fit model
40	0.622	Baik	-0.422	Baik	Fit model

Butir	Daya Beda		Tingkat Kesukaran		Fit Model
41	0.412	Baik	-0.196	Baik	Fit model
42	0.466	Baik	1.271	Baik	Fit model
43	0.585	Baik	0.612	Baik	Fit model
44	0.489	Baik	0.336	Baik	Fit model
45	0.677	Baik	0.173	Baik	Fit model
46	0.323	Tidak Baik	2.897	Tidak Baik	Fit model
47	0.657	Baik	1.865	Baik	Fit model
48	0.356	Tidak Baik	1.598	Baik	Fit model
49	0.442	Baik	1.527	Baik	Fit model
50	0.417	Baik	-1.232	Baik	Fit model

Berdasarkan tabel 3 di atas, diketahui bahwa untuk Tingkat kesukaran butir (*b*) dari 41 butir yang dianalisis, yang dikategorikan butir yang baik sebanyak 33, sedangkan yang dikategorikan butir dengan tingkat kesukaran tidak baik sebanyak 8 butir dan indeks daya beda (*a*) dari 41 butir yang dianalisis, yang dikategorikan butir yang baik sebanyak 30 sedangkan yang dikategorikan indeks daya beda tidak baik sebanyak 11 butir. Kesesuaian butir (fit model) dari 41 butir yang dianalisis, yang dikategorikan butir yang fit dengan model sebanyak 37, sedangkan yang dikategorikan butir tidak fit dengan model sebanyak 4 butir .

Berdasarkan pendekatan Teori Respon Butir, untuk menganalisis karakteristik soal secara empiris digunakan bantuan program Bilog. Hasil analisis program bilog untuk soal UAS Sekolah Dasar Kelas IV Tema 1 di Desa Paramasan Bawah sesuai dengan Tabel 4

Tabel 4. Karakteristik Butir Soal berdasarkan Kriteria Tingat Kesukaran dan Daya beda

Karakteristik Butir	Baik	Tidak Baik
Daya beda	30	11
Tingkat Kesukaran	33	8
Fit terhadap Model	37	4

Berdasarkan tabel 4 menunjukkan bahwa daya beda butir yang baik sebanyak 30 butir soal antara lain 1, 3, 8, 11, 15, 21, 22, 23, 24, 25, 26, 27, 30, 31, 32, 33, 34, 35, 36, 37, 39, 40, 41, 42, 43, 44, 45, 47, 49, 50 sedangkan butir tidak baik sebanyak 11 butir antara lain 5, 6, 9, 10, 14, 16, 18, 19, 28, 46, 48 dari 41 butir yang dianalisis. Tingkat kesukaran butir yang baik sebanyak 33 antarlain 3, 5, 8, 11, 15,18, 21, 22, 23, 24, 25, 26, 27, 28, 30, 31, 32,

33, 34, 35, 36, 37, 39, 40, 41, 42, 43, 44, 45, 47, 48, 49, 50 Sedangkan butir dengan tingkat kesukaran yang tidak baik antara lain 1, 6, 9, 10, 14, 16, 19, 46.

Adapun kemampuan peserta didik berdasarkan analisis menggunakan program Bilog terdapat pada tabel 5.

Tabel 5. Kemampuan Peserta Didik berdasarkan Program Bilog

Peserta Didik	Ability	Skor
Siswa_01	0.3251	56.502
Siswa_02	-0.3692	42.616
Siswa_03	-0.7238	35.524
Siswa_04	-0.1721	46.558
Siswa_05	-0.2734	44.532
Siswa_06	-1.1278	27.444
Siswa_07	-0.3232	43.536
Siswa_08	-0.6979	36.042
Siswa_09	0.1513	53.026
Siswa_10	-0.4624	40.752
Siswa_11	0.4025	58.05
Siswa_12	-0.3972	42.056
Siswa_13	0.416	58.32
Siswa_14	-0.9462	31.076
Siswa_15	-0.5468	39.064
Siswa_16	-0.6931	36.138
Siswa_17	-0.7955	34.09
Siswa_18	1.5663	81.326
Siswa_19	2.3192	96.384
Siswa_20	2.0577	91.154
Siswa_21	1.1257	72.514
Siswa_22	1.2809	75.618
Siswa_23	1.0396	70.792
Siswa_24	0.0228	50.456
Siswa_25	0.2077	54.154
Siswa_26	0.9915	69.83
Siswa_27	0.9971	69.942
Siswa_28	-0.2022	45.956
Siswa_29	0.3123	56.246
Siswa_30	-0.5877	38.246

Peserta Didik	Ability	Skor
Siswa_31	-0.6962	36.076
Siswa_32	-0.4965	40.07
Siswa_33	-0.6487	37.026
Siswa_34	-0.3828	42.344
Siswa_35	-1.0281	29.438
Siswa_36	-0.473	40.54
Siswa_37	-0.2958	44.084

Karakteristik Butir Berdasarkan Teori Tes Klasik

Berdasarkan pendekatan teori tes klasik, untuk menganalisis karakteristik butir soal baik tingkat kesukaran maupun daya pembeda secara empiris digunakan bantuan program *Iteman*. Berdasarkan tingkat kesukaran soal, yang dikategorikan butir soal yang baik adalah rentang antara 0,3 sampai dengan 0,8 dan berdasarkan daya pembeda butir soal yang dapat ditedikategorikan butir soal baik adalah 0,2 sampai dengan +2.

Tabel 6. Karakteristik Butir Soal berdasarkan Kriteria Tingat Kesukaran dan Daya Pembeda

No	Parameter	Butir Baik		Butir Tidak Baik	
		Jumlah	Persen %	Jumlah	Persen %
1	Daya Pembeda	15	30	35	70
2	Tingkat Kesukaran	27	54	23	46

Tabel 6 menunjukkan bahwa soal UAS Kelas IV Tema 1 Sekolah Dasar yang dikategorikan baik menurut tingkat kesukarannya sebanyak 27 butir soal, sedangkan menurut daya pembeda butir soal yang dikategorikan baik sebanyak 15 butir soal dari 50 butir soal yang di analisis.

Adapun indek daya beda butir yang dikategorikan baik antara lain: 3, 8, 22, 23, 24, 26, 27, 30, 31, 32, 40, 43, 44, 45, 47 sedangkan butir yang memiliki indeks daya beda yang dikategorikan tidak baik antara lain: 1, 2, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 25, 28, 29, 33, 34, 35, 36, 37, 38, 39, 41, 42, 46, 48, 49, 50. Tingkat kesukaran butir yang baik sebanyak 27 antarlain 5, 8, 11, 15, 18, 20, 21, 22, 23, 24, 25, 26, 27, 30, 31, 32, 34, 35, 36, 37, 39, 40, 41, 43, 44, 45, 50 Sedangkan butir dengan tingkat kesukaran yang tidak baik antara lain 1, 2, 3, 4, 6, 7, 9, 10, 12, 13, 14, 16, 17, 19, 28, 29, 33, 38, 46, 47, 48, 49.

Adapun kemampuan peserta didik berdasarkan analisis menggunakan program iteman terdapat pada tabel 7.

Tabel 7. Kemampuan Peserta Didik berdasarkan Program Iteman

Peserta Didik	Butir Benar	Skor
siswa_01	18	36
siswa_02	16	32
siswa_03	13	26
siswa_04	15	30
siswa_05	14	28
siswa_06	10	20
siswa_07	16	32
siswa_08	18	36
siswa_09	21	42
siswa_10	15	30
siswa_11	19	38
siswa_12	15	30
siswa_13	18	36
siswa_14	15	30
siswa_15	15	30
siswa_16	11	22
siswa_17	15	30
siswa_18	28	56
siswa_19	32	62
siswa_20	30	60
siswa_21	24	48
siswa_22	25	50
siswa_23	26	52
siswa_24	17	34
siswa_25	16	32
siswa_26	23	26
siswa_27	23	26
siswa_28	16	32
siswa_29	19	38
siswa_30	16	32
siswa_31	16	32
siswa_32	13	26
siswa_33	15	30
siswa_34	15	30
siswa_35	11	22
siswa_36	15	30
siswa_37	15	30

KESIMPULAN

Berdasarkan hasil dan pembahasan diatas, maka dapat disimpulkan:

1. Tingkat kesukaran butir soal yang dikategorikan baik berdasarkan teori respon butir sebanyak, 33 butir soal. Indeks Daya beda yang dikategorikan baik berdasarkan Teori Respon Butir Sebanyak 30 butir soal.
2. Tingkat kesukaran butir soa yang dikategorikan baik berdasarkan teori tes klasik sebanyak 27 butir soal. Indeks Daya beda yang dikategorikan baik berdasarkan Teori Tes Klasik sebanyak 15 Butir.

DAFTAR PUSTAKA

- Allen, M. J., & Yen, W. M. (1979). Introduction to measurement theory. Belmont, CA: Wadsworth. Inc.[Context Link].
- Hambleton, R. K., & Jones, R. W. (1993). An NCME instructional module on: Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38–47.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage.
- Liang, T., Wells, C. S., & Hambleton, R. K. (2014). An assessment of the nonparametric approach for evaluating the fit of item response models. *Journal of Educational Measurement*, 51(1), 1–17.
- Mansyur, S., & Harun, R. (2015). Asesmen pembelajaran di sekolah: Panduan bagi guru dan calon guru. Yogyakarta: Pustaka Pelajar.
- Mardapi, D. (2012). Pengukuran penilaian dan evaluasi pendidikan. Yogyakarta: Nuha Medika.
- Naga, D. S. (1992). Pengantar teori sekor pada pengukuran pendidikan. Jakarta: Gunadarma.
- Retnawati, H. (2014). Teori respons butir dan penerapannya: Untuk peneliti, praktisi pengukuran dan pengujian, mahasiswa pascasarjana. Yogyakarta: Nuha Medika.
- Saifuddin, A. (2002). *TES PRESTASI fungsi dan pengembangan pengukuran prestasi belajar*. Yogyakarta: Pustaka Pelajar Offset.
- Sarea, M. S. (2018). Karakteristik Soal Ujian Akhir Semester Pendidikan Agama Islam Dan Budi Pekerti Tingkat Sekolah Dasar. *An-Nahdhah*, 11(2), 303–318.

- Sarea, M. S., & Hadi, S. (2015). Analisis Kualitas Soal Ujian Akhir Semester Mata Pelajaran Kimia SMA di Kabupaten Gowa. *Jurnal Evaluasi Pendidikan*, 3(1), 35–43.
- Schulz, W., & Fraillon, J. (2011). The analysis of measurement equivalence in international studies using the Rasch model. *Educational Research and Evaluation*, 17(6), 447–464.
- Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (edisi revisi)*. Trim Komunikata Publishing House.
- van der Linden, W. J., & Hambleton, R. K. (2013). *Handbook of modern item response theory*. Springer Science & Business Media.
- Widoyoko, E. P. (2012). Teknik penyusunan instrumen penelitian. *Yogyakarta: Pustaka Pelajar*, 15, 22.